Audio Engineering Society

Convention Paper

Presented at the 130th Convention
2011 May 13–16 London, UK

Spaciousness Rating of 8-channel Stereophony-Based Microphone Arrays

Laurent S. R. Simon[1 2] and Russell Mason[1]

[1] *Institute of Sound Recording, University of Surrey, United Kingdom*

[2] *now with INRIA, centre de Rennes - Bretagne Atlantique, France*

Correspondence should be addressed to Laurent S. R. Simon (laurent.s.simon@inria.fr)

ABSTRACT

In previous studies, the localisation accuracy and the spatial impression of 3-2 stereo microphone arrays were discussed. These showed that 3-2 stereo cannot produce stable images to the side and to the rear of the listener. An octagon loudspeaker array was therefore proposed. Microphone array design for this loudspeaker configuration was studied in terms of localisation accuracy, locatedness and sound image width. This paper describes an experiment conducted to evaluate the spaciousness of 10 different microphone arrays used in different acoustical environments. Spaciousness was analyzed as a function of sound signal, acoustical environment and microphone array's characteristics. It showed that the height of the microphone array and the original acoustical environment are the two variables that have the most influence on the perceived spaciousness, but that microphone directivity and the position of sound sources is also important.

1. BACKGROUND

3-2 stereophony (also known as 5.1) has been shown to produce unstable images, as well as a poor localisation accuracy, to the side and to the rear of the listener [1].

Research has been undertaken to create an 8-loudspeaker array, resulting in a horizontal octagonal arrangement, as shown in fig. 1, and this demonstrates more even localisation around the 360 ° of the horizontal plane compared to a 5.1 ITU-R BS.775-1 loudspeaker array [2]. In order to successfully develop microphone techniques for this reproduction system based on optimum localisation, it was necessary to elicit localisation curves for each of the eight segments of the octagon. The aim of this experiment was to create a map of angle perception and

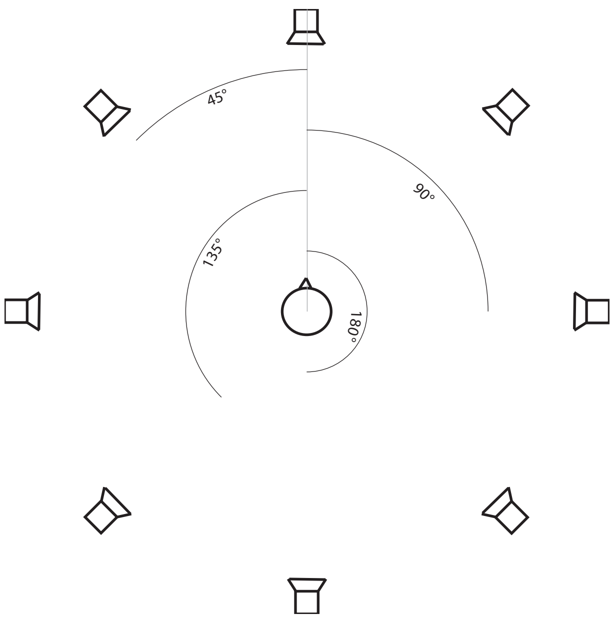

Fig. 1: The octagon loudspeaker setup used in the experiment

locatedness for each segment of the system depending on the interchannel level difference (ICLD) and interchannel time difference (ICTD).

The microphone array were designed for this octagon loudspeaker array on the bases of localisation accuracy. It was shown that the localisation profile of the microphone arrays could be predicted relatively accurately from the localisation curves and an estimation of crosstalk levels and delays. However, other spatial attributes resulting from these 8-channel microphone arrays were not yet investigated as a function of the microphone array's design.

Whilst a microphone array's localisation profile may appear to be an important factor in choosing one microphone array over another, sound engineers sometimes prefer microphone arrays that offer greater spaciousness despite a poor localisation of sound images [1]. In addition, it was shown by Ceoen [3] that near coincident or spaced microphone arrays offer greater spaciousness than coincident microphone techniques. This shows the need for a microphone array design tool that can predict the spaciousness as well as the localisation profile of a microphone array.

Therefore, after studying 8-channel microphone array design in terms of localisation profiles, as reported in [2], the influence of the microphone array design on spaciousness was evaluated. It was expected that the perception of spaciousness would be influenced by the time and level differences between the channels.

The first section of this paper describes the experiment set-up. The the results are then discussed in comparison to previous studies on microphone arrays and spaciousness.

2. EXPERIMENTAL DESIGN

2.1. Selection of experimental conditions

In the same manner as variations in physical parameters of a microphone array, such as the diameter of the array or the directivity of the microphones, caused changes to the resulting localisation profiles, it was expected that these would also alter the perceived spaciousness.

Ten 8-channel microphone arrays were therefore designed and are described in table 1. The different geometries caused changes in terms of maximum InterChannel Time Differences (ICTDs), InterChannel Level Differences (ICLDs)[1], crosstalk levels and crosstalk delays.

Two families of microphone arrays were used: regular and irregular microphone arrays. Regular microphone arrays make the assumption that the listener is free to move his head and might therefore face any direction, whereas the irregular microphone arrays are optimised in terms of localisation profile for a given direction of orientation.

As discussed previously, to the side of the listener, ICTDs lead to less stable phantom images and using solely ICTDs means that the phantom image cannot be positioned inside one of the loudspeakers [4]. Therefore, when designing a microphone array for precise localisation, it is better to use a technique mainly based on ICLDs to the side of the listener and mainly ICTD to the front and rear of the listener, leading to the design shown in fig. 2. An alternative microphone array producing mainly ICLDs to the front and rear of the listener was designed for

[1]Interchannel differences being used in this paper for differences between adjacent microphone channels only.

Type of array	Diameter (m)	Directivity
1	0.5	Cardioid
1	2.5	Cardioid
1	0.5	Omnidirectional
1	2.5	Omnidirectional
1	0.5	Supercardioid
1	2.5	Supercardioid
2	N.A.	Cardioid
2	N.A.	Omnidirectional
3	N.A.	Cardioid
3	N.A.	Omnidirectional

Table 1: List of the microphone array conditions used in the experiment. Type 1 arrays are regular microphone arrays. Type 2 arrays are irregular microphone arrays producing mainly ICLDs to the side of the listener and type 3 microphone arrays are irregular microphone arrays producing mainly ICLDs to the front of the listener. For irregular arrays, the directivity specified is that of the two single microphones; the other microphones in the array are as shown in fig. 2 and 3

comparison purposes, as shown in fig. 3. For these two microphone arrays, the two omnidirectional microphones can be replaced by cardioid microphones in order to reduce the front / back (for the array designed with with mainly side ICLDs) or side (for the array designed with with mainly front / rear ICLDs) crosstalk.

2.2. Original acoustical environments

According to [5] and [6], room reflections can play an important role in perception of space and source localisation. Spaciousness is often thought to be strongly correlated to the room's reverberation time, as a long reverberation time tends to indicate that the original acoustic environment (room) was large.

Each microphone array was therefore used in two different rooms: a small classical studio (room 1), having an RT60 of 1.03s, and a size of approximately 17m width x 14m depth x 7m height and a concert room (room 2), having an RT60 of 2.47s, and a size of approximately 20m width x 30m depth x 20m height.

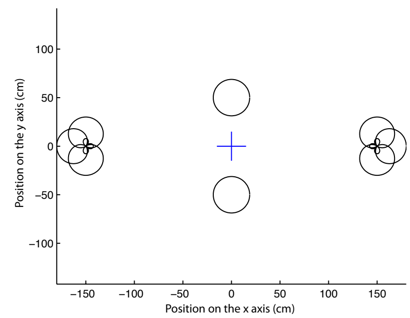

Fig. 2: Microphone array configuration of the side-ICLD microphone array.

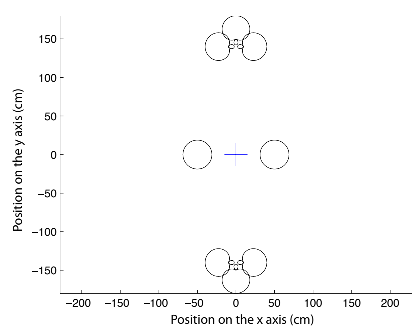

Fig. 3: Microphone array configuration of the front-ICLD microphone array.

2.3. Source positions

10 loudspeakers were positioned around the microphone array, every 36°. The first loudspeaker was located 5° off-centre, in order to avoid having any sound source at the same angle as a loudspeaker. The loudspeakers were located 3 metres away from the centre of the microphone array, as a 6m diameter circle is a likely size for a music ensemble.

The loudspeakers were 10 Genelec 8020As,positioned at a height of 1.35m, which is the expected height for a small musical instrument played in an orchestra by a seated musician.

2.4. Microphone array's height

In order to vary the direct-to-reverberant ratio for a given microphone array, it was necessary to vary the distance of the microphone array to the sound sources. In order to retain the microphone array configuration and to retain the relative angular position of each sound source, the direct-to-reverberant ratio was altered by adjusting the height of the array.

The microphone arrays were therefore used at a height of 1.70m, above the orchestra, as it was the minimum height possible for the microphone structure used to build the microphone arrays, and at a height of 3.75m.

2.5. Recording the audio content

In order to offer the freedom of using any type of source signal, the impulse responses from each loudsepaker to each microphone were captured, and were then convolved with the desired source signals in MATLAB.

The impulse responses were captured on a Macbook Pro laptop using an RME Fireface sound card. Adobe Audition and the Aurora plug-in were used to generate a sine sweep and its inverse response [7]. Then for each microphone array at each microphone height in each room, sine sweeps were played sequentially from each loudspeaker and recorded using all the microphones.

In MATLAB, the impulse responses of each loudspeaker to each microphone of each microphone array were derived, enabling the experimenter to simulate any source signal at any loudspeaker position using any of the microphone arrays under test.

2.6. Sound stimuli

It was thought that the perception of spaciousness would depend on the type of music, on its temporal and spectral characteristics and on the arrangement of the sound images around the listener. Four types of source signal were therefore used:

- A frontal, single female voice.
- A frontal, single drum track.
- A 10-instrument pop music track, each instrument being played by one of the 10 virtual sound sources.
- A 9-instrument classical music track, each instrument being played by one of the 10 virtual sound sources.

2.7. Summary of the recording conditions

The independent variables tested in this experiment were therefore:

- 2 original acoustical environments.
- 2 microphone array heights.
- 4 source signals.
- 10 8-channel microphone arrays, including:
  - 4 irregular microphone arrays.
  - 6 regular microphone arrays, combining 2 different diameters and 3 different directivities.

2.8. Listening test interface

10 trained listeners were asked to rate the spaciousness of the stimuli using a MUSHRA-like interface. The instructions specified that a long reverberation time does not imply that spaciousness is high, in order to limit the bias of misinterpreting the meaning of spaciousness. Spaciousness was rated between not spacious (0% of the scale) and very spacious (100% of the scale). For each source signal, two references were given for the 20% spaciousness and the 80% spaciousness. The references were chosen through an informal test which determined the stimuli thought

to have the lowest and highest spaciousness respectively, though the listeners were informed that they were free to rate the spaciousness higher than the high anchor or lower than the low anchor if they felt it was necessary. In order to evaluate the consistency of the listeners, the two references were also hidden in each set of stimuli.

3. ANALYSIS OF THE RESULTS

In order to check that the data met the assumptions of parametric statistical analysis methods, a Kolmogorov-Smirnov test was carried out for each experimental condition. It showed that all of the cases were normally distributed. This means that the results are suitable for parametric statistical analysis (such as ANOVA) [8].

Several univariate ANOVAs were conducted on the results of the experiment. A first univariate ANOVA was conducted using the type of microphone array, the room, the source signal and the height of the microphone array as dependent variables. It was decided not to include the diameter of the microphone array as it is only relevant to regular microphone arrays. The microphone's directivity was also excluded as it does not apply to the same number of microphones for each type of microphone array (for regular microphone arrays, the microphone's directivity applies to all of the microphones, whereas in the irregular microphone arrays, it only applies to two microphones). This ANOVA showed that the interactions between the source signal and the microphone array height (sig. = 0.000, F = 17.754), the source signal and the room (sig. = 0.000, F = 32.644), the source signal and the type of microphone array (sig. = 0.000, F = 7.646), the type of microphone array and the room (sig. = 0.001, F = 6.847) and between the type of microphone array and the microphone height (sig. = 0.001, F = 6.783) were significant, as can be seen in table 2. While the source signal and the type of microphone array (respectively sig. = 0.003, F = 4.647 and sig. = 0.000, F = 50.947) were found to be significant, the room and the microphone height were found to have a greater F-value by a factor of 5 to 15(respectively sig. = 0.000, F = 256.766 and sig. = 0.000, F = 747.591). According to [8], this means that only the room and the height of the microphone array should be examined further. However, as explained

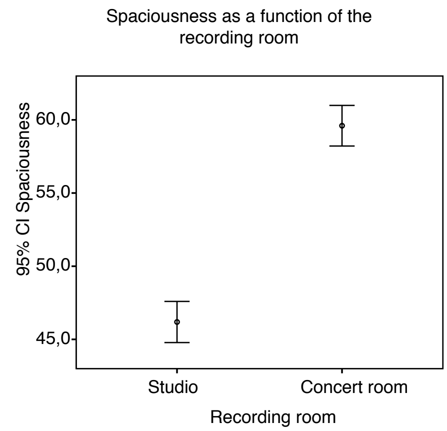

Fig. 4: Spaciousness as a function of the recording room.

in section 1, the spaciousness was expected to be strongly correlated to the reverberation time (hence strongly correlated to the room) and to the direct-to-reverberant ratio (hence strongly correlated to the microphone array height), as can be seen in fig. 4 and 5. While this test confirms this assumption, this paper aims to study the effect of the interaction between these two variables and the microphone array geometry on the perception of spaciousness.

An univariate ANOVA was then conducted for each type of microphone array in order to analyse the influence of the remaining microphone geometry variables on the perception of spaciousness.

For regular microphone arrays, the interactions between the directivity of the microphone, the room, the height of the microphone array and the source signal (sig. = 0.005, F = 3.108), between the diameter of the microphone array, the height of the microphone array and the source signal (sig. = 0.005, F = 3.108) and between the diameter of the microphone array, the height of the microphone array and the directivity of the microphone (sig. = 0.012, F = 4.460) were found significant.

The univariate ANOVA conducted on type 2 microphone arrays showed that for microphone arrays producing mainly ICLDs to the side of the listener, the interaction between the directivity, the height of

Tests of Between-Subjects Effects Dependent Variable:Spaciousness					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	599727,739	47	12760,165	36,204	,000
Intercept	4609833,334	1	4609833,334	13079,156	,000
Type	35913,507	2	17956,754	50,947	,000
Room	90498,941	1	90498,941	256,766	,000
Height	263493,253	1	263493,253	747,591	,000
Signal	4913,312	3	1637,771	4,647	,003
Type * Room	4826,657	2	2413,329	6,847	,001
Type * Height	4781,482	2	2390,741	6,783	,001
Type * Signal	16169,125	6	2694,854	7,646	,000
Room * Height	1283,549	1	1283,549	3,642	,056
Room * Signal	34517,086	3	11505,695	32,644	,000
Height * Signal	18772,545	3	6257,515	17,754	,000
Type * Room * Height	1214,893	2	607,447	1,723	,179
Type * Room * Signal	1962,820	6	327,137	,928	,473
Type * Height * Signal	3427,774	6	571,296	1,621	,137
Room * Height * Signal	726,943	3	242,314	,688	,560
Type * Room * Height * Signal	573,541	6	95,590	,271	,951
Error	772584,607	2192	352,456		
Total	7639587,000	2240			
Corrected Total	1372312,346	2239			
a. R Squared = ,437 (Adjusted R Squared = ,425)					

Table 2: Results of the univariate ANOVA conducted on all types of microphone arrays

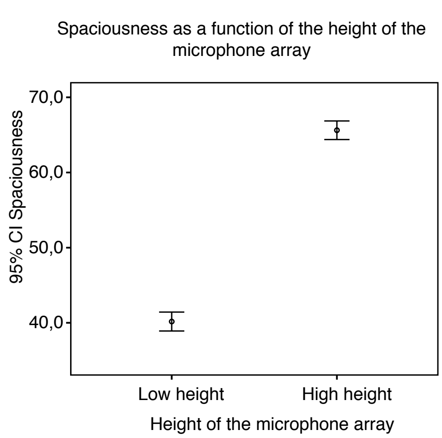

Fig. 5: Spaciousness as a function of the microphone array height.

the microphone array and the source signal (sig. = 0.010, F = 3.861), between the room and the source signal (sig. = 0.000, F = 10.807) and between the room and the height of the microphone array (sig. = 0.000, F = 8.594) were found significant.

Finally, the univariate ANOVA conducted on type 3 microphone arrays showed that for microphone arrays producing mainly ICLDs to the front of the listener, the interaction between the directivity, the room and the source signal (sig. = 0.003, F = 4.644), between the height of the microphone array and the source signal (sig. = 0.000, F = 6.520) and between the directivity and the height of the microphone array (sig. = 0.000, F = 13.486) were found to have a significant influence on the perceived spaciousness.

It can be seen in fig. 6 that the difference of spaciousness for all microphone arrays between the studio room and the concert room is more significant for single source stimuli than for the multiple sound

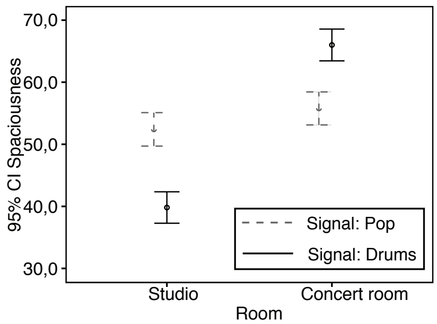

Fig. 6: Spaciousness as a function of the room for both the drums stimuli and the pop music stimuli.

sources stimuli (the figure for the voice stimulus is similar to that of the drum signal and the figure for the classic music stimulus is similar to that of the pop music signal). This could be caused either by the masking of the diffuse field by the sound sources arranged around the listener or because spaciousness could be influenced by the source envelopment. It is also likely that sources with a high source envelopment mask part of the diffuse sound.

Fig. 7 shows that whilst for a low height and a small array diameter, the voice is perceived to be significantly less spacious than the other signals, and that for a low height and a large diameter, the pop signal is perceived significantly more spacious than the other signals, for a high height, all signals are perceived as spacious as the others.

As can be seen in fig. 8, cardioid regular arrays are perceived to be as spacious as supercardioid regular arrays. However, microphone arrays using omnidirectional microphones produce more spacious recordings. This could be explained by the fact that omnidirectional microphones capture more of the diffuse field than cardioid and supercardioid microphones.

Fig. 9 shows that type 2 microphone arrays, (i.e. irregular microphone arrays producing mainly ICLDs to the side of the listener) are perceived less spacious when the two single microphones are omnidirectional microphones and not cardioid microphones, and that as for the rest of this study, the higher microphone arrays are rated as being more spacious than the

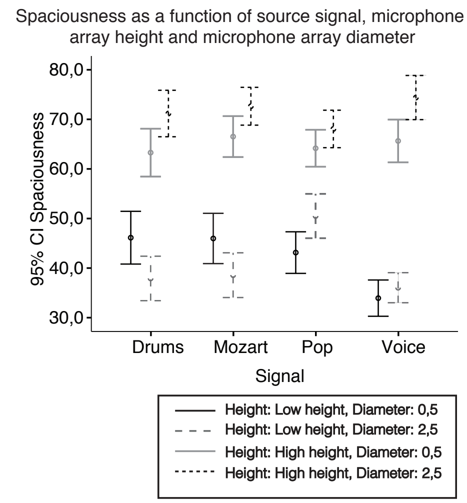

Fig. 7: Spaciousness as a function of the stimuli, microphone array height and microphone array diameter.

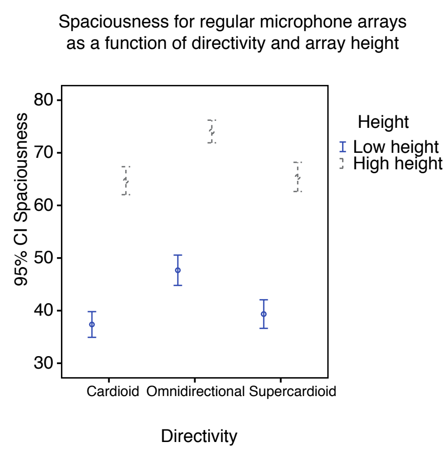

Fig. 8: Spaciousness as a function of the height of the microphone array and directivity of the microphones for regular microphone arrays.

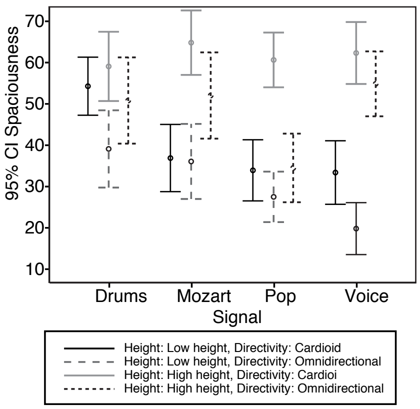

Fig. 9: Spaciousness as a function of the stimuli, microphone array height and directivity of the two single microphones, for irregular microphone arrays producing mainly ICLDs to the side of the listener.

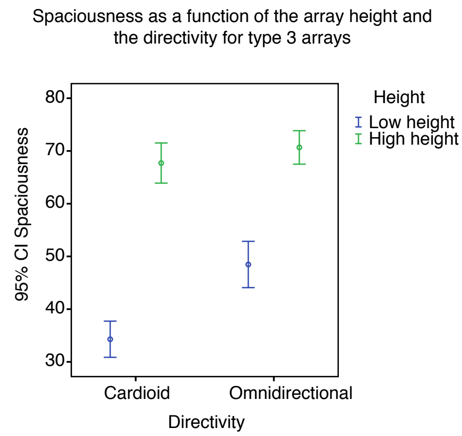

Fig. 10: Spaciousness as a function of the microphone array height and of the two single microphones' directivity, for irregular microphone arrays producing mainly ICLDs to the front and rear of the listener.

lower microphone arrays. It is however unclear why one type of source signal can sometimes be perceived more or less spacious than the three others.

As can be seen in fig. 10, for type 3 arrays, while there is a large difference of spaciousness between the arrays using cardioid microphones and the arrays using omnidirectional microphones at a low microphone array height, the difference is smaller at a high microphone array height.

4. DISCUSSION

It was seen in this experiment that there is a strong correlation between the room in which the recordings were made and the perceived spaciousness of the recording. It was expected that the bigger the recording room, the more spacious the recording.

Similarly, the microphone height appeared to have a large effect on the results, at least as much as the room. For the low microphone arrays, the direct to reverberant ratio is high. The spaciousness of the direct sound might therefore dominate. On the contrary, for the high microphone arrays, the direct to reverberant ratio is low. The spaciousness of the reverberation might then dominate. Hence, for the low microphone arrays the direct sound is dominant, and with a small array diameter there is insufficient decorrelation between the channels. This leaves the spectral characteristics as the main cue to spaciousness, and hence based on research that indicates that low frequencies are important to spaciousness [5], those lacking in low frequencies are perceived to be least spacious. This results in the voice being rated as least spacious. For the low microphone arrays with a large array diameter, there is a greater decorrelation between the channels. In this case, the spatial arrangement of the sources is more important. This results in the pop extract being rated as most spacious. For the high microphone arrays, the reverberant sound is dominant, so the characteristics of the direct sound are less important.

It can also be seen that whilst for type 2 microphone arrays, using omnidirectional microphones causes the recording to be perceived less spacious, using omnidirectional microphones for regular arrays and type 3 arrays causes the recording to be perceived more spacious, as can be seen in fig. 8, fig. 9 and fig. 10. This could be caused by the fact that part of the signals recorded by the two cardioid / omnidirectional microphones, in the case of type 3 or regular microphone arrays, are played by the side loudspeakers. Using cardioid microphones therefore means that the sound played to the side of the listener is less diffuse than when using omnidirectional microphones, and thus perceived as being less spacious, according to Griesinger. However, using omnidirectional microphones for the type 2 microphone arrays is likely to increase front-back confusions [2], which might appear un-natural and cause the listeners to give a lower spaciousness rating to the recordings, despite the absence of a previously known relationship between spaciousness and front-back confusions.

It was also seen that microphone arrays producing only ICLDs to the side of the listener appear less spacious than the two other types of microphone arrays when sound sources are located all around the listener. According to [5], envelopment is caused by the early reflections coming from the side of the listener. Griesinger assumes that spaciousness and envelopment both define the same perceptual phenomenon. Though this is arguable, as the link between the perception of how spacious a recording space sounds (spaciousness) and the feeling of being enveloped by the sound (envelopment) is not intrinsic. However, while the difference between both definitions is clear, the perceptual difference might not be. Some confusion might therefore happen between rating the spaciousness and the envelopment. With a type 2 microphone array, the six side microphones are closer to the side sound sources than with the other two types of microphones. The direct-to-reverberant ratio in the side channels is therefore larger than for the two other types of microphones arrays. This limits the effect of the reflections coming from the side of the listener, thus diminishing the perceived envelopment, and possibly the perceived spaciousness.

Similarly, it was shown that stimuli using multiple sound sources located around the microphone array led to smaller differences of spaciousness between the studio room and the concert room, but that the number of sound sources has little influence on the difference of spaciousness between the high and low height microphone arrays. The number of sound sources was not the only variable: the program material with multiple sources were classical music and pop music whereas the program material with single sources were a drum track and a voice track; the sound level was more constant over time with the multiple sources stimuli than with the single sources stimuli, thus masking more of the reverberation of the room, particularly the end of the reverberation tail. The direct-to-diffuse ratio might have been less affected by this. It could be hypothesised that the presence of virtual sound sources to the side of the listener masked more particularly the side reflections that are important for the perception of envelopment, as discussed above, and that this could cause a smaller difference of spaciousness between the two rooms for multiple source stimuli than for single source stimuli. However, there should then have also been a smaller difference of spaciousness between the two heights for multiple sources stimuli than for single sources stimuli, which was not the case.

5. CONCLUSION

In this paper, it was shown that spaciousness is highly correlated to both the recording room and to the height of the microphone array. It was hypothesised that spaciousness is therefore correlated to the direct-to-reverberant ratio.

Moreover, it was shown that microphone arrays that produce mainly ICTDs to the front and rear of the listener and ICLDs to the side of the listener cause the recordings to be perceived less spacious than the other microphone techniques tested, although this might be caused by the proximity of the side microphones to the side sound sources.

It was also shown than when recording a single frontal sound source, the recording room has a stronger influence on the perceived spaciousness than when recording multiple sound sources around the microphone array.

6. ACKNOWLEDGEMENTS

The research documented in this paper was supported by EPSRC, grant number EP/E064507/1. Its presentation was supported by the EUREKA Eurostars i3Dmusic project funded by Oseo.

7. REFERENCES

[1] Rumsey, F., “Spatial audio”, Focal Press, 2001.

[2] Simon, L. , “Developement of multichannel recording and reproduction techniques based on auditory perceptual cues”, Ph.D. thesis, University of Surrey, 2010.

[3] Ceoen, C., “Comparative Stereophonic Listening Tests”, Journal of the Audio Engineering Society, Vol. 20, issue 1, pp. 19-27, January 1972.

[4] Simon, L., Mason R., , “Time and Level Localisation Curves for a Regularly-Spaced Octagon Loudspeaker Array”, Audio Engineering Society 128th convention, London, United Kingdom, October 2010, Preprint 8079.

[5] Griesinger, D., “Objective Measures of Spaciousness and Envelopment”, Audio Engineering Society 16th International Conference: Spatial Sound Reproduction, April 1999, Rovaniemi, Finland, preprint 16-003.

[6] Hartmann, W., “Localization of a Source of Sound in a Room”, Audio Engineering Society 8th International Conference: the Sound of Audio, May 1990, Washington D.C., United States, preprint 8-006.

[7] Farina, A., “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique”, Audio Engineering Society 108th convention, February 2000, Paris, France, Preprint 5093.

[8] Field, A., “Discovering Statistics Using SPSS”, SAGE Publications LTD, ISBN: 1847879071